Three types of Universals; Three types of Variation

David Adger
Queen Mary University of London

April 30th, 2017

1 Universals

There are a number of subparts of language, thought of as a computational system, which deserve consideration for whether they are universal in some sense. Universal can, of course, be taken in a number of ways, perhaps most obviously 'available to all languages but not in every language' vs. 'in every language'. Let's call these Universal_V and Universal_S. Universal_V is where (part of) variability lives. Universal_S is, I think, restricted to fundamental sructure building processes.

As far as Universal_S goes we have at least the following: the capacity to build hierarchical structures; to associate compositional semantics with these; to create dependencies across these structures which depend on the structure and not on contiguous order; to categorize elements of languages into subclasses; the presence of abstract syntactic elements (e.g. features) associated with no concrete phonology. This is already a lot. There are also more tendentious aspects of these properties: the discreteness of subclasses (denied by those working in Fuzzy Grammar); the structure dependent nature of dependencies (denied by those working with neural nets); the existence of phonologically unexpressed constituents (denied by many, including construction grammarians).

There are broadly two views on what Universals_S are: they are emergent from some minimal UG, or they are emergent from non-language specific properties of the mind, specifically how these interact with language use. For example, Hudson writes:

> syntactic patterns (among others) are learned inductively on the basis of experience, with a great deal of very specific information stored in memory about patterns such as subject-verb pairs (Goldberg (2006)). ... Moreover, in this usage-based account, our memories of tokens may include their contextual specifics, such as who uttered them and when ...

Turning to Universals$_V$, we can think of whether something in one language is available to all as a question about whether that property is acquirable during the normal progress of language development in an individual. Obviously, if some property is acquirable (under normal circumstances) by a speaker of one language, it is acquirable by speakers of all languages. Some of the properties in Universals$_V$ are possibly specific to language, and some are possibly not.

The argument that some are specific to language comes from the a strong case that Universals$_V$ is a subset of all cognitively possible linguistic universals, a fact noted by functional linguists as well as generativists. For example Talmy (1985) remarks

> There are many characteristics of an event's participants that are not marked anywhere in the verb complex, even though they seem as reasonable (from an a priori perspective) as the qualities that are marked. Thus, while an argument's numerosity and distribution can be marked, there will be no marking for its color or whether it has a symmetrical arrangement, even though these very qualities are important in other cognitive systems, such as visual perception? (p.134).

Similarly, many languages have miratives, where a verbal form grammatically expresses that the speaker's attitude to the proposition they are making is one of surprise, but no language has a 'solicitative', marking that the speaker is worried about the situation. Cinque (2013b) provides a range of similar examples and asks the crucial question: "To say that the external and internal temporal constituency of an event (tense and aspect) or the attitude of the speaker toward the truth of the proposition (mood) are cognitively salient is beside the point. The question remains why these and only these cognitive distinctions are encoded grammatically in natural languages out of the many other salient ones." His conclusion is that UG determines what conceptual categories are grammatically encoded. There is a universal inventory: Universals$_V$.

The problems with such a universal inventory was noted decades ago by Susan Steele and others:

> Unless all members of the list of categories in Universal Grammar are found in every natural language, the ?universal inventory? is simply a catalogue of the properties which can be found in natural human languages. The universal inventory cannot be fixed in advance of knowing what the possible categories are; whenever we come upon a language with a new category, even a totally idiosyncratic one, the category can always be added to the list of categories in universal grammar. Thus, in effect, the notion of a fixed universal vocabulary places no limits on the grammars of particular languages, although this is precisely what it is intended to do. Steele et al. (1981), p11.

As well as the theory of what a possible syntactic structure is, we need some kind of a theory of what the content of these is. Cartography and traditional typology provide a data-source for such a theory, but do not, in themselves, provide the theory itself. However, the absences are telling, and more importantly, the interaction of Universals$_V$ and Universals$_S$ leads to a new kind of universal, which it is, I think, possible to theorize fruitfully about.[1]

This third kind of universal is the order of functional categories in an extended projection. Is this an issue of structure or content? In a sense

[1] Universals$_V$ and Universals$_S$ are obviously analogous to Substantive and Formal universals a la Katz and Postal:

> Universals of language are of two different types: substantive universals and formal universals. A linguistic description is a theory and, as such, consists of a set of statements formulated in a fixed theoretical vocabulary. The distinction between substantive universals and formal universals is intended to correspond to the distinction between the form of such statements and their content. Thus a formal universal is a specification of the form of a statement in a linguistic description, while a substantive universal is a concept or set of concepts out of which particular statements in a linguistic description are constructed. (Katz and Postal 1964, p160)

it's neither, but Cartographic research suggests descriptively that there are universals here.

Cinque's view is that the organization of functional categories with respect to each other in an extended projection must be given as an independent syntactic stipulation, disconnected from semantics (Cinque 2013a). I think that this is a currently a minority perspective. Though little is known about how the categories in an extended projection are motivated by conceptual factors, there have been significant proposals. For example, following, ultimately, the lead of Ernst (1998), Ramchand and Svenonius (2014) propose that the reason that C contains T, which in turn contains v, is due to a kind of semantically based mereology of propositions, situations and events. Particular categories in the clause function to transition between these aspects of the semantics of a proposition. In such a proposal the building up of syntactic structure is simultaneously the building up of some aspects of semantic structure.

There are also other ways of conceiving of the reasons for why elements in an extended projection appear where they do, however, that sustain the notion of an autonomous structure building system interacting intimately with systems of thought and meaning, while being quite distinct from them. Take an example like the following:

(1) a. Those three green balls
b. *Those green three balls

As is well known, the order of the demonstrative, numeral and descriptive adjective in a noun phrase follow quite specific typological patterns arguing for a hierarchy where the adjective occurs closest to the noun, the numeral occurs further away and the demonstrative is most distant (Greenberg 1963, Cinque 2005). Why should this be? It seems implausible for this phenomenon to appeal to a mereological emantic structure. I'd like to propose a different way of thinking about thisthat relies on the way that a purely autonomous syntax interfaces with the systems of thought.

Imagine we have a bowl which has red and green ping pong balls in it. Assume a task (a non-linguistic task) which is to identify a particular group of three green balls. Two computations will allow success in this task:

(2) a. select all the green balls
b. take all subsets of three of the output of (a)
c. identify one such subset.

(3) a. take all subsets of three balls
 b. for each subset, select only those that have green balls in them
 c. identify one such subset

Both of these computations achieve the desired result. However, there is clearly a difference in the complexity of each. The second computation requires holding in memory a multidimensional array of all the subsets of three balls, and then computing which of these subsets involve only green balls. The second simply separates out all the green balls, and then takes a much smaller partitioning of these into subsets involving three. So applying the semantic function of colour before that of counting is a less resource intensive computation. Of course, this kind of computation is not specific to colour—the same argument can be made for many of the kinds of properties of items that are encoded by intersective and subsective adjectives.

If such an approach can be generalized, then there is no need to fix the order of adjectival vs. numeral modifiers in the noun phrase as part of an autonomous system. It is the interface between a computational system that delivers a hierarchy, and the use to which that system is put in an independent computational task of identifying referents, plus a principle that favours systems that minimize computation, that leads to the final organization. The syntax reifies the simpler computation via a hierarchy of categories. This means that one need not stipulate the order in UG, nor, in fact, derive the order from the input. The content and hierarchical sequence of the elements in the syntax is delivered by the interface between two distinct systems. This can take place over developmental timescales, and is, of course, likely to be reinforced by the linguistic input, though not determined by it. Orders that are not isomorphic to the easiest computations are allowed by UG, but are pruned away during development because the system ossifies the simpler computation. Such an explanation relies on a generative system that provides the structure which the semantic systems fill with content.

2 Variation

What does it mean for something to be variable? One common-sense view of variability is that a single unit (at some level of abstraction) can come in a variety of forms; for example, pea-plant seeds could vary in whether they are smooth or wrinkled, or clover plants could vary in whether they have three leaves or four.

Within structural linguistics the notion of 'single unit' with a 'variety of forms' was usually conceived of as being connected to the notion of 'linguistic level'. From a classic structuralist perspective (e.g. Harris 1946), each level is an abstraction over some more concrete level. So a phoneme is an abstraction over various phones, a morpheme is an abstraction across morphs, syntactic categories are abstractions over distributional classes, etc.

Schematically, we can write descriptive statements about the kind of variation we find here using rewrite rules sensitive to context:

(4) $$\alpha \rightarrow \begin{cases} a\_\_C_1 \\ A\_\_C_2 \\ \aleph\_\_C_3 \end{cases}$$

Labov's early work (e.g. Labov 1969 et seq.) can be seen as removing the categorical assumption that context *determines* allophony, within this perspective of level-mapping. Labov augmented context sensitive rules with probabilities to capture how the variation is structured by other factors, such as social factors, importing these non-phonological categories into phonological rules. Adger (2006) calls this kind of approach to variation *Variation in Exponence*. This Variation in Exponence approach, modelled by probabalistic context-sensitive rules, lies at the heart of the variationist sociolinguistics enterprise.

At the syntactic and semantic levels, however things become more complex (as is well recognized in the variationist literature, see especially Lavandera 1978, Romaine 1984, Cheshire 1987). The fundamental problem here is how to determine the equivalence between two syntactic forms and a single semantic interpretation, a problem that does not arise at the phonological or morphophonological levels. Even if we assume that semantic representations are individuated simply by truth conditions, the abstract nature of syntactic representations makes the issue somewhat vexed.

To see the issue more clearly, let us look first at the ambiguity case before turning to variation. Take, for example, a quantifier scope ambiguity:

(5) Every leopard chased an owl.

Some theories assume that the syntactic structure simply underdetermines the semantic interpretation, so there is a single syntax mapping to two meanings, much like the case of phonological neutralization (a single glottal stop mapping to two phonemes). For example, Cooper's 1983 approach was to enrich the semantic representations with a 'store' which could be used to disambiguate the sense, and there are also other approaches that take there to be a single syntactic/semantic representation mapping to a multiplicity of interpretations (e.g. Hendriks 1993). In any case, here we have a true one-to-many mapping.

However, a common alternative denies the one-to-many mapping for such cases and takes there to be no ambiguity in the mapping: rather, there is an ambiguity in the assignment of structure to a string or words, of similar sort to that seen in *I saw the boy with the telescope* (e.g. May 1977). If this is the case, then there is actually a one-to-one mapping, rather than a one-to-many mapping, between syntactic structure and semantic representation. For the wide scope reading of the indefinite object, we have a syntax where *an owl* is moved covertly to a higher position than *every leopard* (I represent traces of moved constituents as copies surrounded by angled brackets throughout), while in the narrow scope reading it is raised to a lower position:

(6) a. [an owl] [every leopard] chased ⟨an owl⟩
  b. [every leopard] [an owl] chased ⟨an owl⟩]

Now we have two structures, and each one corresponds to a different semantic reading, so there is a one-to-one mapping between syntactic structures and meanings. This assumption, that the syntactic representations are to a great extent isomorphic on the semantic ones, essentially rules out true ambiguity.

This general viewpoint lends itself to also ruling out variation in exponence, since there's a one-to-one relationship. This has led researchers within generative grammar to treat syntactic variability as involving multiple grammars (Kroch 1989), where each grammar is invariable (see also Roeper 2000, Yang 2002), to build probabilities into the syntactic or morphological systems of the competence grammar, following the work of Labov (1972) and Cedergren and Sankoff (1974) (for more recent attampts to do this, see Bender 2001, Bresnan et al. 2001 and, to a certain extent, Nevins and Parrott 2008). An alternative is to deny that there is real intra-personal variability, and to take the parametric variation between the grammars of individuals in the same community to be of a very subtle form, so that apparent intra-personal variability is to be analysed as being actually inter-personal variation (Henry 2005).

However, there is, within Minimalism, a very straightforward way to create a non-one-to-one syntax-semantic relation. In that theory, syntactic dependencies between positions in structures are encoded by a relation between interpretable and uninterpretable features, usually called Agree. Once an uninterpretable feature is checked by its matching interpretable feature and has done its syntactic work, it is removed from the representation as far as semantic interpretation is concerned. This means that our current theory, as it stands, makes a clear prediction: we should actually find variation within single grammar just when grammatical features enter into agreement or other syntactic dependencies that involve feature checking.

This is the Combinatorial Variability model that is proposed in Adger (2006). The core idea is the following: when two syntactic elements are in an Agree/Checking relationship, one of them will bear uninterpretable features. Since those features are uninterpretable, their presence will not impact on the semantic interpretation of the structure. Given this, a choice of lexical items A and B will be available to agree with C when either of A or B bear uninterpretable features that can match with C, but are distinct from each other. Schematically, we can take C to have three interpretable features f_1, f_2 and f_3:

(7) $C[f_1, f_2, f_3]$

Now if B, which has an interpretable feature g_1, also bears an uninterpretable f_1 feature (uf_1), B can combine with C as follows:

(8) $C[f_1, f_2, f_3]$ $B[g_1, uf_1]$

B's uninterpretable feature will be checked, and unavailable to the semantics. Let's now look at A. In this scenario, A bears the interpretable feature g_1, and an uninterpretable uf_2 feature:

(9) $C[f_1, f_2, f_3]$ $A[g_1, uf_2]$

As far as the semantics is concerned, (9) and (8) are identical, as the uninterpretable features do not feed into the semantic interpretation. However, the syntax and ultimately the systems of spell-out are sensitive to these features, so A and B can have different pronunciations. Further, I show in Adger (2006), that there are cases where we can have yet more lexical items bearing the same interpretable features, but different subsets of uninterpretable ones. Depending on how the phonological form of these features is specified, we can capture different probabilities of particular phonological surface forms. Again, to see this schematically, let us take a third lexical item, D, which also bears interpretable g_1, but in this case uninterpretable f_3:

(10) C[f_1, f_2, f_3] D[g_1, uf_3]

Once again, (10) (that is, CD) has the same semantic interpretation as CB and CA. Now if A, B and D have different phonologies, we will predict that each will have a 0.333 probability of occurrence. However, how do we determine whether we have A, B or D? We have to look at their phonological forms. Obviously, if A, B, and D have three different phonologies, then we will expect each phonological form to appear roughly one third of the time in a large enough corpus. However, imagine that A and D are syncretic: that is, they have the same phonology (call it P1), while B has a different phonology, P2. In that case we will see P1 roughly two thirds of the time, and P1, one third of the time, on the assumption that choice of the three lexical items is random.

This now gives us a system which allows variability within a single grammar (so no multiple grammars). The number of variants is upwards bounded by the number of uninterpretable features on the head though because of the possibility of syncretism, the number of variants may not match the number of uninterpretable features. In addition to this, however, it further provides a mechanism with which to model the frequencies we find in corpora. The mechanism is simply the distribution of syncretisms across the forms in the relevant agreement paradigm. To predict (rather than just model) the frequencies, we further need to specify a formal learning algorithm that will build a lexicon of items which have the appropriate syncretic pattern. I proposed such a learner in Adger (2006), but other kinds of learner are possible (see, for example, Pertsova 2007), and the issue of what kinds of generalization actual learners come up with is, at present, open.

In the Combinatorial Variability model, the grammar (**G**) produces a *Pool of Variants*, PoV, where each variant is a distinct feature complex, with the same semantic interpretation, and with potentially different phonological forms.

(11) $\mathbf{G} \rightarrow \{ v_1, \ldots v_i \ldots v_n \}$ (=PoV)

I assume a distinction between knowledge of language and use of language (Chomsky 1965), so that **G** is embedded in a performance model. One can conceive of the systems of use **U** as a choice function on the pool of variants, given a context of utterance **C**:

(12) $\mathbf{U}(\text{PoV}, \text{C}) = \text{v}_i \in \text{PoV}$

The function **U** is extremely complex, and is sensitive to all sorts of properties of the elements of PoV: their phonology, their sociolinguistic connotations, whether they have been encountered recently, their frequency of occurrence in the life of the language user who is speaking, whether the language user likes that particular word, etc. It is also sensitive to many aspects of the context of utterance: the information structure of the discourse, pragmatic expectations about the interlocutor's knowledge, social expectations about appropriateness etc. Crucially, though, none of these are in the grammar. The probability of any particular v_i being chosen in any speech event is a function of the factors which enter into the specification of **U** and **C**. The fact that some phonological form might be more common in a corpus than some other phonological form depends on both the factors specifying **U** and **C** and the structure of PoV itself. So we must distinguish between the probability of any variant, and the frequency of the particular phonological forms (note how this is subtly different from the classical Labovian notion of linguistic variable).

In this model **G** does not contain sociolinguistic information, unlike the non-modular systems developed by Labov and others. There is never any rule of grammar that makes reference to frequency of a variant, or to social status of a variant. Grammar is sensitive only to syntactic features and structures built up from them. **U**, on the other hand, is part of the performance systems that impacts on the choice of a particular variant in ways that do depend on the speech situation, and the speaker's sociolinguistic capacity; however, **U** does not construct syntactic representations or constrain dependency relations between constituents. Moreover, this general model sees **U** as a dynamically changing function, responsive to the particularities of the utterance situation, and taking into account all of the multifarious factors that influence the particular choice of variant and that are the subject matter of much sociolinguistic research (hence the point about subjective probabilities immediately above).

These then are, plausibly, two different types of variation: Variation in Exponence and Combinatorial Variability. There is, however, I think, a third type of variation too.

Adger (2017) argues that nominal predication in Gaelic and nominal predication in English use quite different means to get to a similar LF. Extending Adger and Ramchand (2003), and following Adger (2013) more generally, I argue that there is no predicational layer in the NPs, so they never have an argument taking capacity (cf. Baker 2003), as true arguments require event identification to be introduced. The category N creates predicates of individuals, not events. This set of constraints on the syntax-semantics interface leaves languages with a problem: how do they build the meaning of NP predication? I argue in that paper that Gaelic shows us two of the ways in which a language can solve this problem. One strategy involves co-opting structure which does have an event variable, in this case from a prepositional aspectual element. This is why we find one variety of nominal predication in Gaelic uses a preposition:

(13) Tha Calum na oileanach.
Be.PRES I in.3S student
'Calum is a student.'

However, Gaelic also uses a different strategy for nominal predication: a Cleft structure combines predication without an NP with arelative clause to create the necessary semantic glue.

(14) 'S e cat a th' ann an Lilly
COP it cat REL be.PRES in Lilly
'Lilly is a cat.'

What of languages like English? As noted in the introduction, nominal predication is restricted in such languages too, when the presence of the verb *be* is controlled for. Nominals are decidedly odd in *be*-less predication compared to PPs and APs:

(15) a. With Lilly *(being) a small cat, she can squeeze thriough the hole.
b. With Lilly sick, we should get some special cat food.
c. With Lilly under anaesthetic, we can go ahead with the operation

From the perspective of the theory offered in this paper, English *be* is performing a function similar to, but more general than, Gaelic *ann an*. Indeed, even with *be*, we can see the same restriction we found in Gaelic, where, when the predicate is restricted to be an interval state by using a temporal modifier, relative clause modification becomes impossible:

(16) ?*Calum was a student for three years that Ian knew.

The same core principles regulating the relationship between syntax and semantics are at work in both kinds of languages, but they evade the restrictions imposed by those principles in different ways. This suggests a third type of variation: cross-linguistic, and indeed intra-linguistic, variation can involve not just setting of parametric choices (say, in terms of feature strength/interpretability), but also how the syntactic resources of a language are deployed combinatorially to create structures which the syntax-semantics interface can interpret in appropriate ways.

References

Adger, David. 2006. Combinatorial variability. *Journal of Linguistics* 42:503–530.

Adger, David. 2013. *A Syntax of Substance*. Cambridge, MA: MIT Press.

Adger, David and Ramchand, Gillian. 2003. Predication and equation. *Linguistic Inquiry* 34:325–359.

Baker, Mark. 2003. *Lexical Categories*. Cambridge, U.K.: Cambridge University Press.

Bender, Emily. 2001. Syntactic variation and linguistic competence. Ph.D. thesis, Stanford, CA.

Bresnan, J., Dingare, S., and Manning, C. D. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In *Proceedings of the LFG 01 Conference*, Stanford: CSLI.

Cedergren, Henrietta and Sankoff, David. 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50:333–355.

Cheshire, Jenny. 1987. Syntactic variation, the linguistic variable and sociolinguistic theory. *Linguistics* 25:257–282.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36:315–332.

Cinque, Guglielmo. 2013a. Cognition, universal grammar, and typological generalizations. *Lingua* 130:50–65.

Cinque, Guglielmo. 2013b. Word order typology: A change in perpective. In Theresa Biberauer and Michelle Sheehan, eds., *Theoretical Approaches to Disharmonic Word Order*, 47–73, Oxford University Press.

Cooper, Robin. 1983. *Quantification and Syntactic Theory*. Dordrecht: Reidel.

Ernst, Thomas. 1998. The scopal basis of adverb licensing. In Pius N. Tamanji and Kiyomi Kusumoto, eds., *Proceedings of the North East Linguistic Society*, 127–142, University of Toronto: Graduate Linguistic Student Association.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg, ed., *Universals of Language*, 73–113, Cambridge: MIT Press.

Harris, Zellig. 1946. From morpheme to utterance. *Language* 22:161–183.

Hendriks, Herman. 1993. Studied flexibility: categories and types in syntax and semantics. Ph.D. thesis, University of Amsterdam., Amsterdam.

Henry, Alison. 2005. Idiolectal variation and syntactic theory. In Leonie Cornips and Karen Corrigan, eds., *Syntax and Variation: Reconciling the Biological and the Social*, 109–122, Amsterdam: John Benjamins.

Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.

Labov, William. 1969. Contraction, deletion and inherent variability of the English copula. *Language* 48:773–818.

Labov, William. 1972. *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.

Lavandera, Beatriz. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7:171–182.

May, R. 1977. The grammar of quantification. Ph.D. thesis, MIT, Cambridge, MA.

Nevins, Andrew and Parrott, Jeffrey K. 2008. Variable rules meet Impoverishment theory. *Lingua* 120:1135–1159.

Pertsova, Katya. 2007. Learning form-meaning mappings in the presence of homonymy. Ph.D. thesis, UCLA.

Ramchand, Gillian and Svenonius, Peter. 2014. Deriving the functional hierarchy. *Language Sciences* 46:152–174.

Roeper, Tom. 2000. Universal bilingualism. *Language and Cognition* 2:169–186.

Romaine, Suzanne. 1984. On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. *Folia Linguistica* 18:409–437.

Steele, Susan, Demers, Richard, Jelinek, Eloise, Kitagawa, Chisato, Oehrle, Richard, and Wasow, Thomas. 1981. *An Encyclopedia of AUX: A Study of Cross-Linguistic Equivalence*. Cambridge, Massachusetts: MIT Press.

Talmy, Leonard. 1985. Lexicalization patterns: semantic structure in lexical forms. In T. Shopen, ed., *Language Typology and Syntactic Description III: Grammatical Categories and the Lexicon*, Cambridge, UK: Cambridge University Press.

Yang, Charles D. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.